Detecting and Correcting Conservativity Principle Violations in Ontology Mappings

Alessandro Solimando

Dipartimento di Informatica, Bioingegneria, Robotica e Ingegneria dei Sistemi,
Università di Genova, Italy
alessandro.solimando@unige.it

1 Problem Statement

Ontologies play a key role in the development of the Semantic Web and are being used in many diverse application domains such as biomedicine and energy industry. An application domain may have been modeled according to different points of view and purposes. This situation usually leads to the development of different ontologies that intuitively overlap, but that use different naming and modeling conventions.

The problem of (semi-)automatically computing mappings between independently developed ontologies is usually referred to as the *ontology matching problem.* A number of sophisticated ontology matching systems have been developed in the last years [5, 30]. These systems, however, rely on lexical and structural heuristics, and the integration of the input ontologies and the mappings may lead to many undesired logical consequences. In [13] three principles were proposed to minimise the number of potentially unintended consequences, namely: *(i) consistency principle*, the mappings should not lead to unsatisfiable classes in the integrated ontology, *(ii) locality principle*, the mappings should link entities that have similar *neighbourhoods*, *(iii) conservativity principle*, the mappings should not introduce new semantic relationships between concepts from one of the input ontologies. Violations to these principles may hinder the usefulness of ontology mappings. Our aim is to develop effective and efficient detection and correction techniques for violations of the conservativity principle for ontology alignments.

2 Relevancy

Given the formal semantics of ontologies, logical defects in the alignment between them may hinder their usefulness and lead to undesired results. The practical effects of these logical violations highly depend on the intrinsic nature and characteristics of the ontology-based system.

When ontology-to-ontology mappings are used, for instance, in an ontology-based data access (OBDA) [27] or an ontology-based data integration system (OBDI) [35], a high quality alignment is mandatory. In such scenarios, any violation of the consistency or conservativity principles will directly affect the quality of the query results, since queries will be rewritten according to the ontology axioms, the ontology-to-ontology mappings and the ontology-to-database mappings. On the contrary, an ontology-based information retrieval (IR) system may better tolerate some logical defects.

A definition of effective techniques for assessing and re-establishing the logical soundness of ontology-to-ontology alignments would be key for any critical ontology-based system using them, directly or indirectly. When such detection and repair techniques are also efficient, ontology matchers may use them for: *(i)* pruning the usually large search space when computing mappings, *(ii)* computing high quality alignments by minimising the number of logical violations. *LogMap* [12] and *AML* [28] ontology matchers, for instance, have already successfully applied these ideas by including detection and repair techniques for the consistency principle in the mapping computation process.

Our work follows a classic approach to ontology alignment debugging, where the repair can only affect the alignment, considering as immutable the matched ontologies [12, 14, 21, 28]. This is not the only possible approach to the problem. The work presented in [10, 17, 18], for instance, considers the violations of the conservativity principle as possible false positives, based on the potential incompleteness of the input ontologies. Hence, the correction strategy may also insert subsumption axioms to the input ontologies, to enrich their concept hierarchies. Authors in [26] also suggest that fixing the input ontologies may be an alternative for mapping removal.

Nonetheless, there are also important application scenarios in which the aligned ontologies have to be considered as not modifiable. One such example is the EU Optique project.[1] Optique aims at facilitating scalable end-user access to big data in the oil and gas industry (based on an OBDA system). Currently, in the Optique use case, the input ontologies are not modifiable. The query formulation ontology is a domain ontology based on the Norwegian Petroleum Directorate (NPD) FactPages[2] [31] and it is currently preferred by Optique end-users to feed the visual query formulation interface [34]. NPD ontology is not intended to be modifiable by end-users, because it includes knowledge already agreed on by the community. The other is a bootstrapped ontology *directly* linked to the information represented in the database.

In general, our approach aims at developing a technique suitable for any ontology-based system, where the used ontologies are not directly controlled by the system and can be only used as they are. For instance, the authors in [20] apply ontology matching in a multi-agent system scenario in order to allow the exchange and extension of ontology-based *action plans* among agents.

3 Related Work

The three principles mentioned in Section 1 have been actively investigated in the last years (*e.g.*, [11, 12, 13, 21, 22, 23, 28]).

---

[1] `http://www.optique-project.eu/`

[2] `http://factpages.npd.no/factpages/`

In particular, the conservativity principle problem, although indirectly, has been actively studied in the literature. Schlobach [29] originally introduced the *assumption of disjointness* to address the repair of ontologies underspecified in terms of negative constraints (disjointness axioms, in particular). A serious obstacle for the practical success of the techniques based on such assumption is the usually prohibitive number of candidate disjointness axioms to be inserted. Meilicke *et al.* [22] applied this assumption in the context of repairing ontology mappings, and limited the number of disjointness axioms to be inserted by using learning techniques [36]. These techniques, however, typically require a manually created training set. In [7] the authors present an interactive system to guide the expert user in the manual enrichment of the ontologies with disjointness axioms. Clearly, this method is not suitable for scenarios in which no user intervention is possible.

Our approach aims at minimising the subset of candidate disjointness axioms that need to be inserted, without compromising the repair effectiveness. However, in order to be applicable to completely automatic repair scenarios, our method needs to work independently from any manual intervention.

Ontology matching systems have also dealt with the conservativity principle in order to improve the precision (w.r.t. a reference mapping set) of the computed mappings. For example, systems such as *ASMOV* [11], *Lily* [37] and *YAM++* [24] have implemented different heuristics to avoid violations of the conservativity principle. Another relevant approach [2] presents a set of sanity checks and best practices when computing ontology mappings. A preliminary analysis shows that the provided heuristics fail at preventing and solving many violations [32, 33].

Unfortunately, for many of the mentioned approaches, the covered ontology fragment is not clear, and their effectiveness can only be experimentally verified, thus limiting the comparability of different contributions. Conservativity principle highly benefited from the definition of formally grounded methods, instead of heuristics approaches, as the results of *OAEI* in the years demonstrate. To this aim, our main goal is to define an elegant way to detect and solve conservativity principle violations by reducing the problem to a consistency principle violation problem, in the Horn propositional fragment. However, in the literature an efficient and automatic technique for enriching ontologies with disjointness axioms, at the basis of the aforementioned reduction, is still missing.

4 Research Questions

After analysing the related work on the subject, we still consider as open the following research questions: *(RQ.i)* Which consequences may have an alignment violating the conservativity principle in different application scenarios? *(RQ.ii)* Is there a relationship between the violations affecting an alignment and its correctness or completeness? *(RQ.iii)* Which algorithms can be used to compute a repair for an alignment violating the conservativity principle? *(RQ.iv)* Which is the trade-off between completeness and runtime for these algorithms? *(RQ.v)* Which are the consequences of applying ontology alignment evolution techniques on an alignment violating the conservativity principle? *(RQ.vi)* How can conservativity principle violations detection and repair support interactive alignment revision?

5 Hypotheses

(H.i) Conservativity principle violations may harm the correctness of ontology-based systems in relevant application scenarios such as OBDA and OBDI (*RQ.i*). *(H.ii)* Conservativity principle is tighly coupled with the notion of conservative extension [16], an extremely challenging decision problem, and would therefore benefit from approximated repair techniques for achieving scalability on reduced DL fragments (*e.g.*, in the $\mathcal{EL}$ family). This principle could be partly reduced to the consistency principle, but a multi-strategy repair is needed to address the uncovered violation kinds (*RQ.iii,RQ.iv*). *(H.iii)* Ontology alignment evolution algorithms usually propagate violations, this could also affect the optimality of the update strategies (*RQ.v*). *(H.iv)* The detection and repair techniques can be coupled with existing user-driven ontology enrichment of negative constraints [7] and ontology revision techniques [25] (*RQ.vi*).

6 Approach

Addressing the conservativity principle violation requires a detection and repair technique. For violation detection, we propose a complete technique, based on an efficient interval labelling schema [1] for the input/aligned ontologies. Given that not all the violations are independent, we plan to provide a discrimination between direct and derived violations, that would rely on a graph representation [32] of the aligned ontology. This refined notion of violation will offer a fine grained violation rate estimation.

For conservativity violations affecting atomic concepts not involved in a subsumption relationship nor sharing any descendant, the problem can be reduced to a consistency repair by inserting a disjointness axiom between the two concepts. A classic approach for debugging ontologies is to compute a repair by computing a (minimal) *hitting set* over the set of justifications [9] (minimal sets of axioms entailing a consequence). Computing all the justifications for a given entailment is a costly reasoning service, and all the scalable debugging algorithms propose approximate repair computations [15, 21, 24, 26]. To address the scalability problem when dealing with large ontologies and mapping sets, our method actually relies on the (Horn) propositional projection of the input ontologies, but does not ensure completeness [33]. We plan to cover expressive fragments such as $\mathcal{EL}$ terminologies (*e.g.*, using the hyper-graph representation of [4]). Currently, we have adapted the infrastructure provided by *LogMap* matcher[12, 14]. However, other mapping repair systems, such as *Alcomo* [21] or *AML* [28], could be considered. Note that, to the best of our knowledge, these mapping repair systems have only focused on solving violations of the consistency principle.

Instead, for the violations affecting concepts involved in a subsumption relationship, the graph representation [32] will again be used, exploiting the property that part of these violations form a cycle (one half represents the previous subsumption relationship, the other one representing the violation). The detection and repair strategies work on the strongly connected components (SCCs) of the graph representation of the aligned ontology, exploting the well-known relation between SCCs and directed cycles. The approximate repair aims at removing all the cycles corresponding to a violation by computing a solution to an ad-hoc variant of the *Feedback Arc Set* problem [6], encoded as a logic program [32].

The idea of enriching the input ontologies with additional disjointness axioms is not new. The novel aspects of our approach are an automatic and efficient identification and addition of a small set of disjoint axioms, using interval indexing. As already discussed in Section 3, another contribution would be the first method addressing the conservativity principle with a theoretical foundation of the concrete ontology fragment covered by both the detection and repair techniques. Another innovative aspect is the combination of graph-theory and logic programming for addressing the violations that cannot be reduced to the consistency problem. Finally, to the best of our knowledge, despite the attention that ontology alignment debugging and ontology alignment evolution [8, 19] techniques have received in the literature, a combined analysis of the possible interrelation between the two fields is still missing.

7 Preliminary Results

Violation Rate: A preliminary analysis [32, 33] suggests that conservativity principle violations not only deeply affect the alignments computed by the top-level ontology matchers, but also widely affect agreed reference alignments, as emerged from the evaluation of the *Ontology Alignment Evaluation Initiative*[3] (*OAEI*) dataset. Also manually curated alignments, such as *UMLS-Metathesaurus* [3] (*UMLS*), a comprehensive effort for integrating biomedical knowledge bases, suffer from these violations.

Conservativity to Consistency Principle Reduction: The repair algorithm of [33] relies on a reduction of the conservativity to the consistency principle, with promising results. We tested the algorithm on the reference alignments of *OAEI 2013*. The complete detection algorithm takes only 275 seconds to process the aligned ontology SNOMED-NCI (the biggest *OAEI*'s test case). Repair efficiency and effectiveness are also promising. Almost all the violations for the five main tracks of the *OAEI 2013* are fully repaired.

Repair Algorithms: For what concerns the two orthogonal repair techniques, the preliminary results show their efficiency and effectiveness in isolation [32, 33]. From the theoretical standpoint, the two techniques address different kinds of conservativity principle violations, but the concrete effect of combining them still needs to be explored in practice. Moreover, their suitability for an automatic use in an ontology matching process has to be experimentally verified.

---

[3] `http://oaei.ontologymatching.org/`

8 Evaluation Plan

The evaluation phase will provide a quantitative measurement of the hypotheses underlying the different aspects of our proposal:

(H.i) will be investigated using the new track of *OAEI 2014*,[4] that will be addressing the problem of ontology alignment for query answering. The effect of the violations and their repair will be tested using the same metrics proposed by the organisers. To this extent, another interesting evaluation is the comparison with alternative approaches (see Section 3) w.r.t. the same task.

(H.ii) has been already successfully addressed in [33]. The preliminary analysis consisted in the detection of the initial number of violations for the *OAEI* reference alignments, the runtime for computing a repair, the size of the repair and the number of unsolved violations. In [32], a similar analysis has been performed on the alignments computed by *OAEI* participants. In addition, an analysis of the repair effect in terms of the completeness and correctness of the alignments has also been conducted.

(H.iii) The state of the art ontology mapping evolution algorithms will be tested using the publicly available snapshots of SNOMED-CT, FMA and NCI ontologies, and UMLS alignments between them (as already done in [8]). The used metric will be the number of violations with and without a repair step on the source (and possibly the target) alignment. The effect in terms of completeness and correctness against the repaired and original reference alignment will be measured using the standard notions of *precision*, *recall* and *f-measure*.

(H.iv) The practical effect of coupling the automatic detection and repair techniques with a user-driven ontology disjointness addition will be conducted by means of a user survey. In addition, we plan to also evaluate completeness and correctness against a manually defined gold-standard using standard IR metrics.

Finally, the acceptable trade-off between the completeness of the detection and repair algorithms and their runtime will be tested by integrating the implemented techniques into an existing ontology matcher.

9 Reflections

Despite the increasing number of contributions addressing ontology alignment debugging, the conservativity principle has received little attention. A possible explanation is that the negative effects of violations to the consistency principle are already evident for any ontology alignment application scenario, and therefore were considered at an earlier stage. Our claim (Section 5) is that conservativity principle would affect more advanced application scenarios, such as OBDA and OBDI. To this aim, in our opinion, it is extremely significative the introduction of a novel track addressing ontology-based query answering in *OAEI 2014*, given the importance of this venue for ontology matching researchers.

Finally, as discussed in Section 7, we have already been able to accumulate encouraging results for what concerns the violation rate, and the efficiency and

---

[4] `http://www.om2014.ontologymatching.org/`

effectiveness of our detection and repair techniques. This constitute a reasonable guarantee for the feasability of our approach.

Acknowledgements. The author thanks Ernesto Jiménez-Ruiz and Giovanna Guerrini for their invaluable help.

References

1. Agrawal, R., Borgida, A., Jagadish, H.V.: Efficient Management of Transitive Relationships in Large Data and Knowledge Bases. In: ACM SIGMOD Conf. on Manag. of Data (1989)
2. Beisswanger, E., Hahn, U., et al.: Towards valid and reusable reference alignments-ten basic quality checks for ontology alignments and their application to three different reference data sets. J. Biomed. Semant. 3(Suppl 1), S4 (2012)
3. Bodenreider, O.: The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Research 32, 267–270 (2004)
4. Ecke, A., Ludwig, M., Walther, D.: The Concept Difference for $\mathcal{EL}$-Terminologies using Hypergraphs. In: DChanges Workshop (2013)
5. Euzenat, J., Meilicke, C., Stuckenschmidt, H., et al.: Ontology Alignment Evaluation Initiative: Six Years of Experience. J. Data Sem. 15, 158–192 (2011)
6. Even, G., Naor, J.S., Schieber, B., Sudan, M.: Approximating Minimum Feedback Sets and Multicuts in Directed Graphs. Algorithmica 20(2), 151–174 (1998)
7. Ferré, S., Rudolph, S.: Advocatus Diaboli - Exploratory Enrichment of Ontologies with Negative Constraints. In: Int'l Conf. on Knowl. Eng. (EKAW). pp. 42–56 (2012)
8. Groß, A., Dos Reis, J.C., Hartung, M., Pruski, C., Rahm, E.: Semi-automatic adaptation of mappings between life science ontologies. In: Data Integration in the Life Sciences. pp. 90–104. Springer (2013)
9. Horridge, M., Parsia, B., Sattler, U.: Laconic and Precise Justifications in OWL. Int'l Sem. Web Conf. (ISWC) pp. 323–338 (2008)
10. Ivanova, V., Lambrix, P.: A Unified Approach for Aligning Taxonomies and Debugging Taxonomies and their Alignments. In: Eur. Sem. Web Conf. (ESWC), pp. 1–15. Springer (2013)
11. Jean-Mary, Y.R., Shironoshita, E.P., Kabuka, M.R.: Ontology Matching With Semantic Verification. J. Web Sem. 7(3), 235–251 (2009)
12. Jiménez-Ruiz, E., Cuenca Grau, B.: LogMap: Logic-based and Scalable Ontology Matching. In: Int'l Sem. Web Conf. (ISWC). pp. 273–288 (2011)
13. Jiménez-Ruiz, E., Cuenca Grau, B., Horrocks, I., Berlanga, R.: Logic-based Assessment of the Compatibility of UMLS Ontology Sources. J. Biomed. Semant. 2(Suppl 1), S2 (2011)
14. Jiménez-Ruiz, E., Cuenca Grau, B., Zhou, Y., Horrocks, I.: Large-scale Interactive Ontology Matching: Algorithms and Implementation. In: Eur. Conf. on Artif. Intell. (ECAI) (2012)
15. Jiménez-Ruiz, E., Meilicke, C., Grau, B.C., Horrocks, I.: Evaluating Mapping Repair Systems with Large Biomedical Ontologies. In: Description Logics. pp. 246–257 (2013)
16. Kontchakov, R., Wolter, F., Zakharyaschev, M.: Can you Tell the Difference Between DL-Lite Ontologies? In: Int'l Conf. on Knowl. Representation and Reasoning (KR) (2008)

17. Lambrix, P., Dragisic, Z., Ivanova, V.: Get My Pizza Right: Repairing Missing Is-a Relations in $\mathcal{ALC}$ Ontologies. In: Semantic Technology, pp. 17–32. Springer (2013)
18. Lambrix, P., Liu, Q.: Debugging the Missing Is-a Structure Within Taxonomies Networked by Partial Reference Alignments. Data Knowl. Eng. (DKE) 86, 179–205 (2013)
19. Martins, H., Silva, N.: A user-driven and a semantic-based ontology mapping evolution approach. In: ICEIS (1). pp. 214–221 (2009)
20. Mascardi, V., Ancona, D., Barbieri, M., Bordini, R.H., Ricci, A.: CooL-AgentSpeak: Endowing AgentSpeak-DL Agents with Plan Exchange and Ontology Services. Web Intelligence and Agent Systems 12(1), 83–107 (2014)
21. Meilicke, C.: Alignments Incoherency in Ontology Matching. Ph.D. thesis, University of Mannheim (2011)
22. Meilicke, C., Völker, J., Stuckenschmidt, H.: Learning Disjointness for Debugging Mappings between Lightweight Ontologies. In: Int'l Conf. on Knowl. Eng. (EKAW). pp. 93–108 (2008)
23. Melnik, S., Garcia-Molina, H., Rahm, E.: Similarity Flooding: A Versatile Graph Matching Algorithm and Its Application to Schema Matching. In: IEEE Int'l Conf. on Data Eng. (2002)
24. Ngo, D., Bellahsene, Z.: YAM++ : A Multi-strategy Based Approach for Ontology Matching Task. In: Int'l Conf. on Knowl. Eng. (EKAW). pp. 421–425 (2012)
25. Nikitina, N., Rudolph, S., Glimm, B.: Interactive ontology revision. Web Semantics: Science, Services and Agents on the World Wide Web 12, 118–130 (2012)
26. Pesquita, C., Faria, D., Santos, E., Couto, F.M.: To repair or not to repair: reconciling correctness and coherence in ontology reference alignments. In: Ontology Matching (OM) (2013)
27. Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M., Rosati, R.: Linking Data to Ontologies. In: J. Data Sem., pp. 133–173. Springer (2008)
28. Santos, E., Faria, D., Pesquita, C., Couto, F.: Ontology Alignment Repair Through Modularization and Confidence-based Heuristics. arXiv:1307.5322 preprint (2013)
29. Schlobach, S.: Debugging and Semantic Clarification by Pinpointing. In: Eur. Sem. Web Conf. (ESWC), pp. 226–240. Springer (2005)
30. Shvaiko, P., Euzenat, J.: Ontology Matching: State of the Art and Future Challenges. IEEE Transactions on Knowl. and Data Eng. (TKDE) (2012)
31. Skjæveland, M.G., et al.: Publishing the Norwegian Petroleum Directorate's FactPages as Semantic Web Data. In: Int'l Sem. Web Conf. (ISWC) (2013)
32. Solimando, A., Guerrini, G.: Coping with Conservativity Principle Violations in Ontology Mappings. Tech. Rep. DIBRIS-TR-14-01, University of Genova (Jan 2014), `ftp://ftp.disi.unige.it/person/SolimandoA/cycleMappingDbgExt.pdf`
33. Solimando, A., Jiménez-Ruiz, E., Guerrini, G.: Detecting and Correcting Conservativity Principle Violations in Ontology-to-Ontology Mappings. Tech. Rep. DIBRIS-TR-14-05, University of Genova (May 2014), `ftp://ftp.disi.unige.it/person/SolimandoA/conservLogMap.pdf`
34. Soylu, A., et al.: A Preliminary Approach on Ontology-Based Visual Query Formulation for Big Data. In: 7th Research Conf. on Metadata and Semantics Research (MTSR) (2013)
35. Tian, A., Sequeda, J., Miranker, D.P.: QODI: Query as Context in Automatic Data Integration. In: Int'l Sem. Web Conf. (ISWC). pp. 624–639 (2013)
36. Völker, J., Vrandecic, D., Sure, Y., Hotho, A.: Learning Disjointness. In: Eur. Sem. Web Conf. (ESWC). pp. 175–189 (2007)
37. Wang, P., Xu, B.: Debugging Ontology Mappings: A Static Approach. Computing and Informatics 27(1), 21–36 (2012)